

APPARATUS AND METHOD OF CLASSIFYING MOVEMENT OF OBJECTS IN A MONITORING ZONE

INTRODUCTION AND BACKGROUND

5 This invention relates to apparatus and a method of monitoring movement of objects in a monitoring region.

In US 5,519,784 there is disclosed apparatus and a method of classifying movement of objects along a passage. The apparatus
10 comprises means for projecting an array of discrete and spaced parallel linear radiation beams from one side of the passage to an opposite side of the passage. Detectors at the opposite side of the passage sense when the beams are interrupted by one or more persons moving in the passage in either of a first and a second opposite direction. The
15 spaced beams are interrupted at different times in a sequence corresponding to the number of and direction of movement of persons. The sequentially generated interrupted beam signals are stored as object movement historic information in memory and then processed to generate composite beam interrupt patterns manifesting the number
20 of persons and direction of movement, the patterns being a function of time domain and sensor index, i.e. sensor identity and position in the passage. The resulting generated patterns are compared to reference patterns utilizing computerized pattern recognition analysis, such as with an artificial neural network. The comparison classifies the persons
25 in the passage into direction of movement and number.

This apparatus and method may not be suitable for some applications. For example, the means for projecting the spaced parallel beams is mounted in an elongate housing. The housing is normally mounted on one side of the passage to extend parallel to the floor at about
5 between ankle and knee height. This housing may be too large to fit into available space therefor and/or may not be aesthetically acceptable in certain applications. Furthermore, it is labour, time and cost intensive to mount this apparatus on either side of the passage and it often is necessary to chase into the side-walls of the passage to
10 install cabling extending to the apparatus. Still furthermore, the beam projecting means on the one side and the detectors on the other side may become misaligned, which would cause the apparatus to cease functioning. Another problem with this side-on mounted apparatus is that a person or object stationary in the passage could interrupt the
15 beams and hence cause at least temporary insensitivity of the apparatus to other objects moving along the passage on either side of the stationary object. Still a further problem is that the range of the projected beams may not be sufficient to traverse a wide passage. Intermediate structures carrying additional apparatus with all of the
20 aforementioned disadvantages are required to cover the wide passage.

Another known system, but which is fundamentally different, uses tracking algorithms and attempts to identify discrete objects and monitor their position between successive frames produced by an object sensing arrangement, in order to determine a vector for each object of interest. The processing is complex, as it requires a full analysis of each frame and then a comparison to previous frames to determine whether an object is either a previous object in a new position or a new object altogether. Tied in with this, is the difficulty in distinguishing between two people on the one hand and one person carrying a backpack or luggage, for example, on the other. By isolating objects and obtaining their vectors, the system is able to track their movement across a predetermined monitoring region and thus increment or decrement counts accordingly. Any inability of the system to isolate objects, link their successive positions or distinguish number of objects compromises the accuracy of the system. In addition, the visual analysis is extremely processor-intensive and thus expensive.

OBJECT OF THE INVENTION

Accordingly, it is an object of the present invention to provide alternative apparatus and a method of monitoring movement of objects through a monitoring region.

SUMMARY OF THE INVENTION

According to the invention there is provided apparatus for monitoring movement of objects through a monitoring region, the apparatus comprising:

- 5 - a sensing arrangement sensitive to the presence or absence of an object in each of a plurality of adjacent zones in the region;
- the zones being arranged such that there are at least two adjacent rows of zones in a first direction and at least
10 two adjacent rows of zones in a direction perpendicular to the first direction;
- each zone being associated with a respective zone index;
- the sensing arrangement being configured to be sensitive to the presence or absence of an object in each of the
15 zones individually and being operative to capture time sequential representations of objects moving through the region comprising sensed data relating to the presence or absence of objects in each of the zones;
- a processor arrangement connected to the sensing
20 arrangement for processing the sensed data into a multidimensional pattern of the presence or absence of

the objects in the zones, wherein a first dimension is time and a second dimension is zone index;

- the processor arrangement being configured to segment the pattern into pattern parts relating to events; and
- 5 - a classifier for classifying the pattern parts with reference to historical data relating to anticipated events.

The sensing arrangement may comprise at least one camera, which is mounted overhead the monitoring region.

10

The sensing arrangement may comprise a stereo pair of cameras covering the region from different angles.

15

Hence, the system according to invention does not attempt to identify each unique object in the field of view but analyses an event and by comparing this to previous knowledge of event features it is able to give a count figure using the classifier, which may comprise a neural network. The image processing is relatively simple and may be done relatively cheaply, while the neural network itself can run on a relatively simple microprocessor, thus saving on cost. It is believed

20 that the system may also alleviate at least some of the aforementioned

problems associated with the aforementioned side-on mounted system of US 5,519,784.

5 The plurality of zones may form an array of immediately adjacent zones and each zone may have a first dimension in the first direction, a second dimension in the second direction and an area.

10 The sensed data may comprise data or a parameter proportional to a part of the area of the zone being occupied by the object.

The processor arrangement may be configured to segment the pattern along the time dimension, in regions of the pattern of inactivity.

15 According to another aspect of the invention there is provided a method of monitoring movement of objects through a monitoring region, the method comprising the steps of:

- dividing the region into a plurality of zones, each zone being associated with a zone index;
- 20 - the zones being arranged such that there are at least two adjacent rows of zones in a first direction and at least

two adjacent rows of zones in a direction perpendicular to the first direction;

- 5 - utilizing a sensing arrangement automatically and time sequentially to sense the presence or absence of objects in each of the zones individually;
- generating data relating to a multi-dimensional pattern of the presence or absence of the objects in the zones, wherein a first dimension is time and a second dimension is zone index;
- 10 - segmenting the pattern into pattern parts relating to events; and
- classifying the pattern parts with reference to historical data relating to anticipated events.

15 Yet further included in the scope of the present invention is a computer readable medium hosting a computer program for monitoring movement of objects through a monitoring region, the program executing the steps of:

- 20 - receiving from a sensor arrangement sensed data relating to time sequential representations of objects moving through the region, the sensed data comprising data relating to the presence or absence of objects in each of a

5 plurality of zones, the zones being arranged such that there are at least two adjacent rows of zones in a first direction and at least two adjacent rows of zones in a direction perpendicular to the first direction and wherein each zone is associated with a zone index;

- generating data relating to a multi-dimensional pattern of the presence or absence of the objects in the zones, wherein a first dimension is time and a second dimension is zone index;
- 10 - segmenting the pattern into pattern parts relating to events; and
- classifying the pattern parts with reference to historical data relating to anticipated events,

15 thereby, in use, to provide as an output, a count of objects moving in at least one direction through the region.

The invention also extends to firmware comprising a computer program and to a computer program configured to perform the aforementioned steps.

20 **BRIEF DESCRIPTION OF THE ACCOMPANYING DIAGRAMS**

The invention will now further be described, by way of example only, with reference to the accompanying diagrams wherein:

- figure 1 is a diagrammatic representation of a monitoring region with a camera of apparatus for monitoring movement of objects through the region, mounted above the region;
- figure 2 is a diagrammatic representation of an array of zones in the monitoring region;
- figure 3 is a block diagram of the apparatus;
- figure 4 is a flow diagram of relevant parts of a method of monitoring movement of objects through a monitoring region;
- figure 5 is a diagram illustrating time sequential images of an elliptical object moving through the region;
- figure 6 is a three dimensional representation, one dimension being time, of the object moving through the region;
- figure 7 is an alternative representation wherein the three dimensional tensor has been flattened into a two-dimensional matrix;
- figure 8 is a resulting representation of the event of figure 5;
- figure 9 is an image of the event in figures 5 and 9;
- figure 10 is a photograph from above of a plurality of people moving through the region;
- figure 11 is a representation similar to figure 6 of the event in figure 10;

- figure 12 is an image similar to figure 9 of the event in figure 10;
- figure 13 is a graph of distributions associated with a pixel channel;
- figure 14 is a representation of a simple sub event model;
- figure 15 depicts sub-event search images;
- 5 figure 16 is a representation of maxima defining centres of sub-events;
- figure 17 is a representation of the first four dimensions of a sub-event model;
- figure 18 is a representation of sub-events extracted from the event
- 10 in figure 15 and their classifications; and
- figure 19 is a representation of the original event in figure 15 and a residual event after the sub-events in figure 18 have been removed.

15 **DESCRIPTION OF A PREFERRED EMBODIMENT OF THE INVENTION**

Apparatus for monitoring movement of objects through a monitoring region 12 is generally designated by reference numeral 10 in figures 1 and 2.

20 The region 12 may form part of a portal or passage 14 at a counting point, such as an entrance 16 to a building 18 and the apparatus 10 may be deployed automatically and over a period of time to monitor

and count people 26 entering and leaving the building through that entrance, as will hereinafter be described.

The apparatus 10 comprises a sensing arrangement 20 sensitive to the presence or absence of an object 26 in each of a plurality of adjacent zones 24.1 to 24.n in the region. Referring to figure 2, the zones 24.1 to 24.n are arranged such that there are at least two adjacent rows of zones in a first direction $\underline{y}$ (that is a general direction of flow of objects through the region) and at least two adjacent rows of zones in a direction $\underline{x}$ perpendicular to the first direction. Each zone 24.1 to 24.n is associated with a respective zone index. The sensor arrangement 20 is configured to be sensitive to the presence or absence of the object in each of the zones individually and being operative to capture time sequential representations, preferably images (as shown in figures 5 and 6) of an object moving through the region. The images comprise sensed data comprising data relating to the presence or absence of the object in each of the zones. A processor arrangement 22 (shown in figure 3) is connected to the sensing arrangement 20 for processing the sensed data into a multidimensional pattern (shown in figures 7 and 8) of the presence or absence of the object in the zones as the object moves through the region, wherein a first dimension is time and a second dimension is zone index. The processor arrangement is

further configured to segment the pattern into pattern parts relating to events and a classifier for classifying in real time the pattern parts with reference to historical data relating to anticipated events.

5 The sensing arrangement may comprise at least one image sensor, such as a video camera 20 and associated optics 21, mounted at the zone 12, for capturing time sequential images of the zone, each image comprising sensed data. The apparatus further comprises an electronic subsystem 23 comprising the processor arrangement 22 (shown in
10 figure 3) connected to the camera 20 for receiving the sensed data from the camera and for generating the multi-dimensional pattern data as will hereinafter be described.

15 The camera 20 is preferably mounted overhead the passage in a roof 28 and hence above the monitoring region 12. The camera may comprise a stereo pair camera comprising first and second cameras directed at the region at different angles so that they cover the region from different angles to add an extra dimension, and to define the monitoring region 12 at a suitable level or height h above a floor 30 of
20 the passage 14. The subsystem 23 may be provided at the monitoring zone, alternatively centrally at the same building to be connected to similar sensing arrangement at other entrances (not shown) of the

building, further alternatively the subsystem may be positioned remotely and off-site.

5 In an example embodiment shown in figure 2, the monitoring region 12 comprises a first axis $\underline{y}$ extending in a first direction and a second axis $\underline{x}$ extending in a second direction, which is perpendicular to the first direction, so that each zone has an area. The aforementioned plurality of zones 24.1 to 24.n in the monitoring zone 12 form an array of rows and columns of adjacent regions, each having a first
10 dimension in the first direction and a second dimension in the second direction.

In an example embodiment shown in figure 3, the camera 20 is connected to the processor arrangement 22 of the subsystem 23. The
15 processor arrangement comprises a field programmable gate array (FPGA), alternatively a complex programmable logic device (CPLD) 32. The FPGA is connected to a dual port RAM arrangement 34 and the arrangement 34 is connected to a processing core 36. In other embodiments (not shown) the processor arrangement 22 may
20 comprise a camera interface and a processor with Direct Memory Access (DMA). The subsystem 23 further comprises a communications module 38 connected to the processing core 36 and

to logic circuitry 40. Also connected to the logic circuitry are a real time clock (RTC) 42, data memory arrangement 44 and program memory arrangement 46. The apparatus 10 is provided with electrical power, by power supply module 48, which is connectable to mains power. The communications module 38 enables data communications between the apparatus 10 and an external network.

As stated hereinbefore, the camera 20 is configured to capture time sequential images of the region 12. Hence, each image and its associated sensed data are associated with unique time related data, thereby providing a time dimension for the sensed data and a multi-dimensional representation, as will hereinafter be described.

Referring to figure 5, time sequential images generated by the sensing arrangement 20 when an elliptical object 26 moves through the region 12 are illustrated in figures 5(a) to 5(f). For example, at the time of the second image, that is the image in figure 5(b), the object covers zone 24.3 completely and zones 24,2, 24.4 and 24.8 partially. The sensed data may, as a parameter, include data proportional to a part of the area of the zone being occupied or covered by the object.

The stream of sensed data or matrixes in figures 5(a) to 5(f) may be represented in a three-dimensional representation, wherein time is one dimension and as illustrated in figure 6.

5 The three-dimensional representation may be rearranged by vectorizing the matrixes and so flatten the three-dimensional tensor into a two-dimensional representation wherein one dimension is time (t) as illustrated in figure 7.

10 The stream of pattern matrixes is segmented in time, corresponding to periods of activity. An event is triggered whenever the sum of pattern matrix elements exceeds a threshold level over a small set of consecutive frames. Conversely an event is terminated whenever the sum falls below the threshold over a set of consecutive frames.

15

Representations of the event in figure 5 are shown in figures 8 and 9.

In figure 10 there is shown another example event wherein a plurality of people are moving in either direction through the region 12. Figure

20 11 corresponds to figure 6 for the latter event and figure 12 corresponds to figure 9 for this latter event.

Referring now to figure 4, the apparatus and method have the following main components, preprocessing 60, event extraction 70, feature extraction 80 and classification 90 to transform and process a mass of pixel related data in a video volume down to two numbers, namely count in direction A and count in direction B.

The front end video pre-processing 60 is concerned with processing the raw pixel stream into a form that is distinctive yet invariant with respect to the background as well as global variation in brightness and contrast. This part of the system may be the most computationally sensitive as it operates on a per pixel level. Thus, video processing is intentionally kept relatively simple. At 62, the image is first reduced to a manageable size by filtering and down sampling. From this resized image an active region is extracted. The active region is a user specified area in a frame that defines the boundary across which people are counted. For computational efficiency this boundary is required to be straight and image axis aligned. This boundary defines two directions, the first direction of flow y and the second direction x perpendicular thereto. The active region is an axis aligned rectangle that is centred on the boundary. The rectangle's lateral width is specified by the user and its dimension along the direction of flow is determined by the relative scale of an average human in the frame.

Only the active region is processed further and the rest of the frame is discarded. An important task of the pre-processing is to normalise the raw input video with respect to the background. The background in this context shall be defined as the part of the image that corresponds to objects in the scene that are physically static over an extended period of time. The foreground, conversely, corresponds to the parts of the image that depict physically moving objects. To segment each frame as such, models of the pixel variation associated with both the foreground and background are constructed. Since the background is defined by its slow rate of change a background model 64 is approximated by essentially applying a temporal low-pass filter to statistics associated with the input video. A foreground model is constructed by analysing the statistics of image regions not sufficiently described by the background model. This normalisation process is referred to as background removal. It ultimately attempts to assign, to each pixel of an input frame, a probability that it is part of the foreground (a moving object).

For computational simplicity each input pixel is considered independently, each pixel in turn has a set of channels associated with it. The variation within these channels is modelled by a multivariate Gaussian distribution. This choice is weakly motivated by the ubiquity

of the Gaussian distribution due to the Central Limit Theorem, but more so by the fact that the Gaussian distribution can be fitted to input data simply by calculating its average and spread. The multivariate Gaussian distribution of a d -dimensional random variable $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ is as follows:

$$G_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

Often the logarithm of this distribution is more convenient for computation:

$$\ln(G_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\boldsymbol{\Delta})) = -\frac{1}{2} \left[d \ln(2\pi) + \ln |\boldsymbol{\Sigma}| + \boldsymbol{\Delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Delta} \right]$$

Where $\boldsymbol{\Delta} \equiv \mathbf{x} - \boldsymbol{\mu}$

Each pixel is represented by a d dimensional vector $\mathbf{x}$ of pixel channels. Currently four channels are used, the luminance and two chrominance values from the YUV colour space as well as the time derivative of the luminance. Pixel foreground and background conditional distributions as follows:

$$P(\mathbf{x} | \mathbf{x} \in S_{bg}) = G_{\boldsymbol{\mu}_{bg}, \boldsymbol{\Sigma}_{bg}}(\mathbf{x}), \quad P(\mathbf{x} | \mathbf{x} \in S_{fg}) = G_{\boldsymbol{\mu}_{fg}, \boldsymbol{\Sigma}_{fg}}(\mathbf{x})$$

Where S_{fg} and $S_{bg} = S_{fg}^C$ represent the sets of the possible $\mathbf{x}$ that corresponds to the foreground and background respectively and

$\{\mu_{bg}, \Sigma_{bg}\}$ and $\{\mu_{fg}, \Sigma_{fg}\}$ correspond to the mean and covariance associated with the foreground and background respectively. To keep this derivation concise the fg and bg subscripts that denote the two distributions shall be omitted in equations that hold for both the foreground and background models.

For the sake of computational simplicity the pixel channels are assumed independent. Thus Σ is assumed diagonal and so the Gaussian distributions may be expressed as follows

$$\ln[G_{\mu, \Sigma}(\mathbf{x})] = -\frac{1}{2} \left[d \ln(2\pi) + \sum_{i=1}^d \left(\frac{(x_i - \mu_i)^2}{\sigma_i^2} + \ln(\sigma_i^2) \right) \right]$$

Where σ_i^2 correspond to the diagonal elements of Σ and x_i and μ_i are the elements of a $\mathbf{x}$ and μ respectively for $i=1..d$. Given the prior probabilities for the two classes, $P(\mathbf{x} \in S_{fg}) \equiv \gamma_{fg}$, $P(\mathbf{x} \in S_{bg}) \equiv \gamma_{bg}$ the conditional distributions can be transformed into joint distributions:

$$P(\mathbf{x}, \mathbf{x} \in S) = P(\mathbf{x} \in S) P(\mathbf{x} | \mathbf{x} \in S) \equiv p(\mathbf{x})$$

Note the priors are constrained by $\gamma_{fg} + \gamma_{bg} = 1$, since a pixel belongs to either the foreground or background.

The ultimate goal is the posterior probability $P(x \in S_{fg} | x) \equiv z_{fg}(x)$, the probability that a given pixel is part of the foreground. This may be calculated according to Bayes' theorem as follows:

$$z_{fg}(x) = \frac{p_{fg}(x)}{p_{bg}(x) + p_{fg}(x)} = 1 - z_{bg}(x)$$

5

In terms of the logarithmic forms of the distributions, which better represents the actual method of calculation, may be expressed as follows:

$$L(x) \equiv \ln p(x) = \ln \gamma - \frac{1}{2} \left[d \ln(2\pi) + \sum_{i=1}^d \left(\frac{(x_i - \mu_i)^2}{\sigma_i^2} + \ln(\sigma_i^2) \right) \right]$$

10

$$z_{fg}(x) = \left(1 + \frac{p_{fg}(x)}{p_{bg}(x)} \right)^{-1} = \frac{1}{1 + \exp(L_{fg}(x) - L_{bg}(x))}$$

15

The parameters of these distributions, $\{\mu_{bg}, \Sigma_{bg}\}$ and $\{\mu_{fg}, \Sigma_{fg}\}$ are adapted over time to track changing illumination and changes in the background. The background mean μ_{bg} is modelled per pixel however the variances Σ_{fg} , Σ_{bg} and the foreground mean μ_{bg} are global and shared by all pixels in the image. This choice was made to keep the computational complexity down but also to keep the calculation of the variances more stable by averaging the statistic over the entire frame.

Figure 13 is a typical example of the distributions associated with a pixel channel. The posterior probability is close to 1 where the foreground is dominant, but decreases to almost 0 where the background is dominant. The pixel model parameters are gradually updated after each frame so as to track slow changes in statistics caused by variation in global illumination or additions to the background. The model parameters are updated in a manner similar to a first order low-pass IIR filter. Given the current parameters θ_i , the new parameters θ_{i+1} are defined as follows:

$$\theta_{i+1} = \frac{\theta_i + \lambda \hat{\theta}}{1 + \lambda}$$

where λ is the adaptation rate and $\hat{\theta}$ is the vector of parameters obtained by fitting the model to the data in the current frame. The foreground and background means and variances are approximated as weighted first and second moments of the pixel channel values. The pixel weightings are dependent on their associated probabilities. The foreground probability is used in the calculation of the foreground model parameters and the background probability for the background parameters. A much slower adaptation rate is used for the variance calculation since it requires more degrees of freedom, thus more data is needed to reliably determine it. A non-linearity is added to this linear approach by making λ dependent on the foreground and background

pixel probabilities. This is done with the variance update by modulating λ by the fraction of foreground pixels in the current frame. Thus variances only change when there is activity within the frame. This prevents the variances from becoming artificially small over long periods of no activity. This also makes sense in that one is only interested in distinguishing foreground and background at times of activity within the frame.

Once the input video is normalised into a form that is independent of the background and illumination, the video it is broken down into manageable chunks. This is done by down sampling the normalised video into patterns and then segmenting these patterns in time to form events.

Patterns are extracted at 72 for each frame directly from its foreground probability image. The pattern is simply constructed by averaging the foreground probability within the zones 24.1 to 24.n of the grid or array spanning the region. As shown in figures 2 and 5, the grid has four divisions in the direction of flow. The number of divisions in the lateral direction is dependent on the physical width of the active region. The zones' aspect ratio is kept constant, however, their scale is matched to the average relative size of a human in the frame. Thus,

each frame yields a pattern that may be represented by a matrix whose elements correspond to the fraction of foreground pixels in their associated zone.

5 The pattern matrices are stacked over time into what may be thought of as a three-dimensional tensor and as shown in figures 6 and 11. At 74 in figure 4, this stream of pattern matrices is segmented in time into events, corresponding to periods of activity. An event is triggered whenever the sum of pattern matrix elements exceeds a threshold
10 over a small set of consecutive frames. Conversely an event is terminated when the sum falls below the threshold over a set of consecutive frames. The representation in figure 11 depicts each zone as a cube with a size proportional to the zone's value. To compactly visualise these events it is useful to flatten the three dimensional
15 structure along the direction of flow to form a matrix, as illustrated in figure 7 which may be conveniently depicted as a two-dimensional image, as is the case figures 8, 9 and 12. Referring again to figure 4, in an attempt to compact event patterns in time and to some extent normalise the event with respect to object velocity, an approach
20 analogous to Run Length Encoding (RLE) is applied at 76 to the events. The process begins with the first frame pattern as a prototype. Frames are then compared in succession with the prototype, if the

frame is not represented well by the prototype (i.e. the difference is large) it then becomes the new prototype. This process continues until the end of the event is reached. The normalised event is constructed by averaging the stretches between prototype changes into single patterns. The length of each span is also recorded and appended onto each prototype vector so original timing is not lost.

A normalised mean square distance measure is used to compare patterns

$$D(\mathbf{p}_0, \mathbf{p}_1) = \frac{|\mathbf{p}_1 - \mathbf{p}_0|^2}{|\mathbf{p}_1|^2 + |\mathbf{p}_0|^2 + c}$$

Where $\mathbf{p}_i$ are column vectors representing patterns and c is a small positive regularising constant to prevent division by zero. For clarity, any events mentioned in the remainder of this description shall be assumed normalised in this respect.

Feature extraction at 80 in figure 4 generally involves projecting from some usually high dimensional input space associated with an input sensor arrangement 20 to a low dimensional parameter space associated with a model of the objects of interest. In this case, feature

extraction is concerned with projecting a variable length event pattern into a distinctive fixed length feature vector.

Events are further decomposed into regions that correspond to common lower level sub-events such as single people, 2 people close together or a person with a trolley. To do this, a model 82 of such sub-events is constructed. Linear models provide the simplest option for this as there exists closed form solutions to their construction, such as Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA). While these methods produce models that compactly and distinctly represent the sub-events, they do not provide a direct method of classifying them. For this, a Gaussian Mixture Model (GMM) is used to partition the subspace produced by PCA and LDA into classes. The sub-event model consists of two related models, a simple search model used to efficiently find (at 84 in figure 84) sub-events within an event; and a more complex classification model used to provisionally or weakly classify them at 86. Both of these models consist of linear basis constructed using LDA and a GMM defined in the sub space spanned by the basis. Sub-events span a fixed size window within the larger event. In this case a $6 \times 4 \times 16$ window is used. If the window extends beyond the dimensions of the event, zeros are substituted for the out of bounds elements. Augmented to

the sub-event data extracted from within the window, is the relative lateral location of the window as well as the set of spans generated by the RLE compression that correspond to the window frames.

5 As stated hereinbefore, a simple sub-event model is used to efficiently find at 84 valid sub-events within an event. This simple model attempts to distinguish between three classes:

Inward - sub-events that are on average in the inward direction as illustrated at A in figure 2;

10 Outward - sub events that are on average in the outward direction as illustrated at B in figure 2; and

Null - Regions of the event that do not correspond to moving people.

15 The exemplars for each of the classes when projected back into the event space are shown in the following table:

Table 1 Simple sub-event classes

ID	1	2	3
Type	null	in	out
Exemplar			

In figure 14 there is depicted the GMM used to distinguish these classes within the LDM sub space. In the plot, the lines correspond to the iso-probability lines of the individual Gaussian components. Note the Gaussian components have unrestricted covariance matrices, thus the iso-probability lines map to ellipsoids. The background image in the plot depicts the soft decision boundaries generated by the posterior probability of the classes.

Sub-events are found at 84 by exhaustively searching the event in time and along the direction lateral movement. A three-dimensional window, the size of the sub-event model, is slid over the event, classifying each point as either the centre of an inward, outward or null sub-events. This produces a pair of two-dimensional search images as shown in figure 15, corresponding to the inward and outward directions. These images are added together and smoothed to produce an image with well defined maxima as shown in figure 16. It is these maxima that define the centres of the sub-events. The maxima in figure 16 are indicated by black pixels 100 in the blob centres. The inward and outward images are used again later in the in the feature calculation.

Once sub-event centres are found as aforesaid, they need to be weakly classified at 86 in figure 4. This is done with a more elaborate classifying sub-event model. It is more computationally expensive and more discriminating, but is only applied to already found sub-events and not the entire event. This model uses a five-dimensional basis.

Twelve sub-event classes are distinguished by this model. These classes are summarised in the following table:

Table 2 : Sub event classes

ID		1	2	3	4	5	6	7	8	9	10	11	12
People	in	0	1	0	1	0	1	2	0	3	0	2	1
	out	0	0	1	0	1	1	0	2	0	3	1	2
Trolley	in	0	0	0	1	0	0	0	0	0	0	0	0
	out	0	0	0	0	1	0	0	0	0	0	0	0
Exemplar													

Like the simple model each class is modelled as a Gaussian component with unrestricted covariance. Figure 17 shows the first four dimensions of the sub-event model.

The x's in figure 17 plot the sub-events found in the example event in figure 15 with each square labelled according to a rounded version of the sub-event model's classification of it. The corresponding appearance of these sub events in the event space and their classification is shown in figure 18.

The classification sub-event model seems to have done quite well in this particular case as the actual count for the entire event is 4-2. Each sub-event is classified in turn each producing a vector of posterior probabilities where the i^{th} element corresponds to the probability that the sub-event represents the i^{th} class. These vectors are summed over all the sub-events to produce a vector z that forms part of the final feature vector. The components of z roughly correspond to the number of sub-events of each class within the event. If two sub-events are found close together with an appreciable overlap in their associated windows, the possibility arise that the overlapping information is counted twice, once for each sub-event. To mitigate this after each sub-event is classified, the result of the classification is projected back into the event space and subtracted from the original event effectively marking that part of the event as counted. Figure 19 shows the original event and what remains after the sub-events have been successively removed.

The construction of the sub-event model requires that the locations and labels of the sub-events within the training data are known. However, this data is not available, since the training data is not segmented down to sub-event level, there is but a single label per event. Thus, an iterative approach is used to build the sub-event model. The process begins with an initial model constructed under the assumption of a single sub-event per event, centred at the event's centroid. Using this rough model a new set of sub-events are found and classified from which the model may be recalculated and so the process continues until the model converges.

The feature vector $\mathbf{f}$ consists of a set of aggregate statistics over the entire event. The structure of the feature vector is as follows

$$\mathbf{f} = \left[t_e \quad t_c \quad \mathbf{m}_{in}^T \quad \mathbf{m}_{out}^T \quad s \quad \mathbf{z}^T \right]^T$$

Where:

t_e Duration of the event in frames

t_c Length of the compressed event

$\mathbf{m}_{in}$ Moments of the inward search image

$\mathbf{m}_{out}$ Moments of the outward search image

s The sum of the maxima values corresponding to sub events centres

z The total sub event classification vector

The moment vectors $\mathbf{m}_{in}$ and $\mathbf{m}_{out}$ consist of the 0th and 2nd degree moments of the in and out search images illustrated in figure 15. Each moment vector has 4 dimensions of the following form

$$\mathbf{m} = [m_{0,0} \quad m_{2,0} \quad m_{0,2} \quad m_{1,1}]^T$$

Where if $f_{x,t}$ represents an image element at lateral position x and time t

$$\begin{aligned} m_{0,0} &= \sum_x \sum_t f_{x,t} \\ \begin{bmatrix} m_{1,0} \\ m_{0,1} \end{bmatrix} &= \frac{1}{m_{0,0}} \sum_x \sum_t f_{x,t} \begin{bmatrix} x \\ t \end{bmatrix} \\ \begin{bmatrix} m_{2,0} & m_{1,1} \\ m_{1,1} & m_{0,2} \end{bmatrix} &= \frac{1}{m_{0,0}} \sum_x \sum_t f_{x,t} \begin{bmatrix} x \\ t \end{bmatrix} \begin{bmatrix} x & t \end{bmatrix} - \begin{bmatrix} m_{1,0} \\ m_{0,1} \end{bmatrix} \begin{bmatrix} m_{1,0} & m_{0,1} \end{bmatrix} \end{aligned}$$

The sub-event classification vector is twelve-dimensional, as there are twelve sub-event classes, the moment vectors contribute four components each and there are the three scalars, thus the final feature vector is twenty-three dimensional.

Referring again to figure 4, the goal of the classification at 90 is to extract the final "in" and "out"-counts from each feature vector. A mapping between feature vectors and counts needs to be constructed,

which constitutes a regression problem. An artificial neural network 92 is used to implement this mapping. The neural network learns this mapping from a large set of tagged events by optimising the regression error. A Standard Multi-layer Perception (MLP) is used, with
5 a single hidden layer using sigmoid activation functions and output layer with linear activation for regression.

The training of the MLP is essentially an optimisation problem, minimising the output error with respect to the edge weights, for this
10 a conjugate gradient descent algorithm is used. The training data takes the form of a set of feature vectors and corresponding count labels. However, before training, the features are whitened, normalising the feature space to unit covariance, so improving the chance of convergence to an absolute minima. The normalizing projections are
15 incorporated back into the first layer weights of the neural network after it is trained.

It will be appreciated that the output of the neural net could be used to count people moving through the region and any one of the
20 direction A and direction B shown in figure 2.

CLAIMS

1. Apparatus for monitoring movement of objects through a monitoring region, the apparatus comprising:

- 5 - a sensing arrangement sensitive to the presence or absence of an object in each of a plurality of adjacent zones in the region individually;
- 10 - the zones being arranged such that there are at least two adjacent rows of zones in a first direction and at least two adjacent rows of zones in a direction perpendicular to the first direction;
- each zone being associated with a respective zone index;
- 15 - the sensing arrangement being operative to capture time sequential representations of objects moving through the region comprising sensed data relating to the presence or absence of the objects in each of the zones;
- 20 - a processor arrangement connected to the sensing arrangement and for processing the sensed data into a multidimensional pattern of the presence or absence of the objects in the zones, wherein a first dimension is time and a second dimension is zone index;
- the processor arrangement being configured to segment the pattern into pattern parts relating to events; and

- a classifier for classifying the pattern parts with reference to historical data relating to anticipated events.

2. Apparatus as claimed in claim 1 wherein the sensing
5 arrangement comprises at least one camera, which is mounted overhead the monitoring region.
3. Apparatus as claimed in claim 1 or claim 2 wherein the sensing
10 arrangement comprises a stereo pair of cameras covering the region from different angles.
4. Apparatus as claimed in any one of claims 1 to 3 wherein the
15 plurality of zones form an array of immediately adjacent zones and wherein each zone has a first dimension in the first direction, a second dimension in the second direction and an area.
5. Apparatus as claimed in claim 4 wherein the sensed data
20 comprises a parameter proportional to a part of the area of the zone being occupied by the object.

6. Apparatus as claimed in any one of claims 1 to 5 wherein the processor is configured to segment the pattern along the time dimension, in regions of the pattern of inactivity.

- 5 7. A method of monitoring movement of objects through a monitoring region, the method comprising the steps of:
- dividing the region into a plurality of zones, each zone being associated with a zone index;
 - the zones being arranged such that there are at least two
10 adjacent rows of zones in a first direction and at least two adjacent rows of zones in a direction perpendicular to the first direction;
 - utilizing a sensor arrangement automatically and time sequentially to sense the presence or absence of objects
15 in each of the zones individually;
 - generating data relating to a multi-dimensional pattern of the presence or absence of the objects in the zones, wherein a first dimension is time and a second dimension is zone index;
 - segmenting the pattern into pattern parts relating to
20 events; and

- classifying the pattern parts with reference to historical data relating to anticipated events.

5 8. A computer readable medium hosting a computer program for monitoring movement of objects through a monitoring region, the program executing the steps of:

- receiving from a sensor arrangement sensed data relating to time sequential representations of objects moving through the region, the sensed data comprising data relating to the presence or absence of the objects in each of a plurality of zones, the zones being arranged such that there are at least two adjacent rows of zones in a first direction and at least two adjacent rows of zones in a direction perpendicular to the first direction and wherein each zone is associated with a zone index;
 - generating data relating to a multi-dimensional pattern of the presence or absence of the objects in the zones, wherein a first dimension is time and a second dimension is zone index;
 - segmenting the pattern into pattern parts relating to events; and
- 10
15
20

- classifying the pattern parts with reference to historical data relating to anticipated events,
thereby in use to provide as an output, a count of objects moving in at least one direction through the region.

5

9. Firmware hosting a computer program for monitoring movement of objects through a monitoring region, the program executing the steps of:

10

- receiving from a sensor arrangement sensed data relating to time sequential representations of objects moving through the region, the sensed data comprising data relating to the presence or absence of the objects in each of a plurality of zones, the zones being arranged such that there are at least two adjacent rows of zones in a first direction and at least two adjacent rows of zones in a direction perpendicular to the first direction and wherein each zone is associated with a zone index;

15

- generating data relating to a multi-dimensional pattern of the presence or absence of the objects in the zones, wherein a first dimension is time and a second dimension is zone index;

20

- segmenting the pattern into pattern parts relating to events; and
- classifying the pattern parts with reference to historical data relating to anticipated events,

5 thereby in use to provide as an output, a count of objects moving in at least one direction through the region.

10. A computer program for monitoring movement of objects through a monitoring region, the program being configured to
10 execute the steps of:

- receiving from a sensor arrangement sensed data relating to time sequential representations of objects moving through the region, the sensed data comprising data relating to the presence or absence of the objects in each
15 of a plurality of zones, the zones being arranged such that there are at least two adjacent rows of zones in a first direction and at least two adjacent rows of zones in a direction perpendicular to the first direction and wherein each zone is associated with a zone index;
- generating data relating to a multi-dimensional pattern of
20 the presence or absence of the objects in the zones,

wherein a first dimension is time and a second dimension is zone index;

- segmenting the pattern into pattern parts relating to events; and
 - 5 - classifying the pattern parts with reference to historical data relating to anticipated events,
- thereby in use to provide as an output, a count of objects moving in at least one direction through the region.